How Automatic is Adults' Theory of Mind Reasoning?

TANG, Ki Yuen

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Philosophy in

Psychology

The Chinese University of Hong Kong

September 2010

Abstract

In the current literature, there are mixed evidences regarding the automaticity of theory of mind. Three studies were conducted to further investigate the question of whether adults do automatically take into account others' mental states in understanding behavior, and under what circumstances would the deployment of theory of mind be enhanced.

In study 1, replication of Wertz and German's (2007) study on belief-desire reasoning was conducted to collect qualitative feedback on improving the experimental design. Study 2 and study 3 were extension of the original Wertz and German's (2007) study. Through presenting different mixes of false belief and true belief stories to different groups of participants, study 2 provided evidence that people's performance on belief-desire reasoning would be affected by the predominant mental state that the environment called for in general, instead of what was particularly entailed in a single incident. Study 3 showed that highlighting contextual cues related to mental states, such as deception and initial desire, could reduce the error in making mental-states attribution for others' behaviour. The results of these studies are consistent with the argument that theory of mind is strategic in nature and requires devotion of mental resources, but inconsistent with the Theory of Mind Mechanism model arguing for an automatic theory of mind mechanism.

Thesis/Assessment Committee

Professor Takeshi Hamamura (Chair)
Professor Him Cheung (Thesis Supervisor)
Professor Darius Kwan-shing Chan (Committee Member)
Professor Kevin Kien-hoa Chung (External Examiner)

摘要

在現時的文獻中，有關心智理論的自動性有著不同的說法和證據。爲了探討這個問題，我們進行了三個實驗，以進一步理解成年人在理解他人的行爲時，會否自動地顧及他人的心智狀態；與及在甚麼情況下這樣的心智推理表現會有所提升。

第一個實驗重覆了 Wertz 和 German 在二零零七年的信念/慾望推理實驗，以收集意見，從而對實驗的設計進行改良。實驗二和實驗三乃 Wertz 和 German 二零零七年實驗之延續。在實驗二中，不同實驗群組的參與者會看到不同數目和比例的真實信念和錯誤信念故事，令他們傾向用某一種心智狀態 (信念或慾望) 來理解他人的行爲。實驗參與者在信念/慾望推理中的表現會被當時主要使用的心智狀態影響。

實驗三亦進一步證明有關心智狀態的描述 (如把事情描述爲欺騙行爲，或說明故事主人翁原本的慾望) 能減少實驗參與者在信念/慾望推理上的錯誤。這一系列實驗顯示了心智理論的不自動性，亦跟 Leslie, Friedman 和 German 在二零零四年提出的完整的心理機制 (Theory of Mind Mechanism) 不一致。

Introduction

Theory of mind refers to an understanding of mental states such as belief, desire, intention and knowledge, as well as the capacity to attribute these mental states to oneself and others. It enables us to perceive others as distinct entities, providing a way to explain and predict others' behavior through inferring their mental states (Astington, 2003). In order to attribute mental states to others and use them to make sense of behavior, people must be able to appreciate that others generate their own independent mental representations of the world using the information they have, and such representations may differ from one another and reality. This also means that one must be able to maintain different representations of the world simultaneously. It is a 'theory' of mind in that such representations are not 'directly observable'. (Premack & Woodruff, 1978).

Given that the human mind and its thinking are complicated, and a wide variety of information is relevant in understanding and predicting other's action, theory of mind entails understanding of various mental states. In particular, belief and desire are acknowledged to be the most important mental states for the explanation and prediction of human action (Davidson, 1963; Wellman, 1991). People typically engage in actions because they believe those actions will satisfy their desires. On the other hand, we understand others' actions by inferring their beliefs and desires. For example, if one believes that an object is in a particular location and s/he desires the object, s/he is likely to perform an action to retrieve the object from that location. In same vein, if we know that a person did a particular action like searching for an object in a specific location, we can infer the reasons that may have caused this action - s/he desires the object and believes that the object is in that location. Such inferences help

us organize and make sense of the actions performed by the people around us, thus facilitating social interaction.

Theory of mind has long been viewed as a milestone in children's cognitive development. As a milestone, the developmental investigation of theory of mind is often framed as a question of presence or absence, such that people are interested in when and under what circumstances children demonstrate their capacity of understanding others' mental states. What happens beyond the development of such rudimentary theory of mind, however, has largely been overlooked. Few studies have been conducted to investigate how the mature theory of mind works in enabling our understanding and prediction of other people's behavior, or enhancing interaction in the social context. There has not been a comprehensive account on how adults deploy their theory of mind capacity in social interactions, and this clearly is a gap in the current literature that deserves more attention. Indeed, perspective taking, a component of theory of mind that enables us to consider the world in another individual's viewpoint (Galinsky, Maddux, Gilin, & White, 2008), has been shown to be pivotal in a variety of social and interpersonal exchanges, including social cooperation (Paal & Bereczkei, 2007), negotiation (Galinsky et al., 2008), poker game (Lopes, 1976), politics, and business (Findler, 1990; Axelrod, 1987). Given the role of theory of mind in social interaction, it would be beneficial to identify the processes involved such that ways of enhancing smoother social interaction could be found.

As social beings, we readily utilize our theory of mind capability in everyday life, such that we can understand others' action and even predict what they are going to do next. This helps organize information from our surrounding in an ordered and predictable way. With the frequent and seemingly effortless deployment of theory of mind in our personal experience, several theorists have argued that theory-of-mind

must be automatic (Friedman & Leslie, 2004; Sperber & Wilson, 2002), which depends on domain-specific modular processes that are fast, spontaneous, encapsulated, with its functioning largely independent of intellectual general capacities of the individual. However, some recent research has provided evidence otherwise (Apperly, Riggs, Simpson, Chiavarino, & Samson, 2006; Keysar, Lin, & Barr, 2003), suggesting that theory of mind might be strategic in nature, requiring the devotion of attention and cognitive effort. The present study aims at investigating theory of mind in adults, particularly at testing whether theory of mind reasoning is automatic. If not, what are the factors that trigger its deployment? We believe that understanding the cognitive mechanism involved in the performance of a mature theory of mind would not only contribute to a more comprehensive understanding of this important capability, but also shed light on how to improve people's social interaction through its effective deployment.

Literature review

The developing theory of mind

Research on theory of mind has devoted a great deal of attention in pinpointing its developmental path. Successful performance on false belief tasks has long been viewed as a milestone in children's cognitive development, and much effort has been put into identifying the earliest age at which it is demonstrated in children. Owing to wide variation in linguistic and general cognitive ability of children in different age groups, vastly different methods have been devised to tap the understanding of mental states in different target age-groups, ranging from the habituation procedure in infants to story-based scenario studies in young children.

Understanding of intention seems to come first in the developmental pathway. Using a habituation procedure, Gergely, Nadasdy, Csibra and Biro (1995) found that 12-month-old children were able to demonstrate an understanding that intentional agents act in rational ways. In a study on 18 month olds' ability to understand the intentions of others, Meltzoff (1995) found that children mimic intentional, but not unintentional, behaviors of adults in their environment, and that they imitate considerably less often when a machine is performing the behavior. This experiment suggests that infants younger than two years of age may be considering the intentions of others and interpret humans, not machines, as intentional beings.

As children grow up, they began to grasp the concept of desire and the link between desire and action. By age 3, children understand some aspects of the links between people's thinking or feeling and their behavior. In a study done by Lillard and Flavell (1992), it was demonstrated that children of age 3 knew that a person who wanted something would take action to get it, and also that a person might still want something even if s/he could not have it. Yet, they might not understand the basic principle that actions were based on one's own representation of reality, and that a person's representation may differ from the external reality. This was evidenced by the fact that children do not pass the false belief tests reliably until age 6.

The understanding of belief, acknowledging that people could have beliefs that differ from the external reality or from others', is a significant milestone that marks the development of a mature theory of mind. By 6 years old, children could reliably pass false-belief tasks. They have a sophisticated theory of mind that enables them to distinguish between their own mental states and reality, and those of others, as well as understand the actions of others in terms of their underlying mental states.

To summarize, before 2 years of age, infants have some rudimentary understanding of the mental state of intention, viewing people as volitional agents having distinct goals and intentions. From 2 to 3 years, they start to understand the mental state of desire, and its role in guiding people's action. From 4 years onwards, children begin to grasp the mental state of belief, understanding that different people might hold different representations of the world that may or may not concur with objective reality, such that they could reliably pass false belief and other theory of mind tasks.

While the mainstream view is that children's performance on standard theory of mind tasks truly reflects their theory of mind capacity, there has been evidence suggesting that children's task performance may sometimes be significantly affected by extraneous or contextual factors. A number of researchers have claimed that the original false-belief tasks are unnecessarily difficult, confounding false belief understanding with other general cognitive demands, and that children's performance can be improved to above-chance level if the tasks are suitably revised (Cassidy, 1998; Siegal & Beattie, 1991; Sullivan & Winner, 1993). For example, in Sullivan & Winner's (1993) study, when the false-belief task was framed in terms of trickery, children's attribution of false-belief and ignorance improved significantly such that young 3-years olds were able to pass the test. These results have indicated that children of 3 years old or younger can distinguish other's mental states from their own, and that false belief competence is expected to be high among young children if it is assessed correctly (Chandler, Fritz, & Hala, 1989; Fodor, 1992).

Mechanism underlying theory of mind

The investigation into the developmental route of theory of mind leads to the formulation of models underlying this complex cognitive capability. The vast differences in the levels of theory-of-mind development have lent support to the view that the early capacity for mental state representation has an innate, automatic basis. Leslie, Friedman and German (2004) proposed a two-component model of theory of mind, which depicts an innate, automatic cognitive module responsible for ToM that is activated at around 2 to 3 years of age. In their theory of mind mechanism model (ToMM), the first component is a neurocognitive mechanism that takes the behavior of social agents as input information to generate candidate representations that might have contributed to this behavior. After the candidates are generated, a second mechanism called the selection processor chooses the most plausible candidate through inhibition, taking into account whether the belief is true or false as well as whether the desire is to approach or to avoid. In the case of standard false-belief scenario in Wimmer and Perner (1983), ToMM spontaneously identifies at least two possible contents, that is, the object is in the original location (false-belief content) and the object is in the new location (true belief content). These contents are then reviewed by the selection processor, with the true belief content being initially more salient than the false belief content. Yet as it is depicted that the agent has a false belief regarding the location of the target object, the true-belief content is inhibited and rejected such that the false-belief content would be selected as the final candidate. This two-component theory of mind model postulates that theory of mind is based on domain-specific cognitive mechanism, and the gradual maturation of this ToMM is what underlies the observed developmental pathway in theory of mind.

Empirical research has lent support to this view of domain-specificity underlying theory of mind. Using a scenario study, Kinderman, Dunbar, and Bentall

(1998) showed that individual differences in mind-reading ability could not be attributed to variances in working memory capacity, pointing to the conclusion that specific cognitive skills are required for understanding the mental states of a person, which are different from those required for recalling factual events. Evidence from studies on psychiatric disorders such as autism has also been in line with this model. Autistic children perform significantly more poorly on false belief tasks compared to other cognitive tasks testing intelligence and language capacities (Baron-Cohen, 1995; Frith, 1994), which possibly suggests that autism could be one of the consequences of a specific deficit of the domain-specific ToMM. Also, neuroimaging studies on adult participants also showed that specific brain regions are involved in theory of mind performance (Frith & Frith, 1999; German et al, 2004; Saxe et al, 2004), supporting this modular view of theory of mind. In the functional MRI study conducted by German et al in 2004, increased activity was found in specific brain regions, including medial prefrontal areas, when participants viewed video clips of actors performing pretend actions as compared with real actions. This finding supported the automatic and domain-specific aspects of the ToMM model, as it showed that specific brain areas associated with making explicit mental state judgments are also activated in response to actions that call for mental state interpretation, even when the participants have not been explicitly instructed to interpret the mental states of the actor.

A study done by Wertz and German (2007) also argued for the automaticity of the ToMM mechanism. In their study, participants were required to read short stories similar to the change-in-location task, in which an actor places a target object in one location, leaves the room, and then another character moves the target object to a second location. To retrieve the target object, the first actor approaches the initial location on her return, where a distracter object is either present or absent depending

on the experimental condition. It was found that when the story characters were described as searching in the wrong place for a target object, resulting in their approach to the distracter object, adult subjects were more likely to explain the character's action as desiring the distracter object. Through this study, the authors proposed that when people are presented with information about the actions of a social agent, cognitive mechanism automatically generates candidate belief-desire explanations for that action. In the cases where the action of an agent is directed towards a particular object, mental states related to that object, including a desire for it and a true belief about it, are represented and influence decisions about what likely caused the observed action. In other words, when we see others approaching an object, we would have a tendency to reason that they approach that object because they want the object and they know the object is there, even if the original scenario depicts that the character approaches that location because s/he has a false belief about another target object.

While ToMM depicts theory of mind as an innate, hard-wired cognitive module, the theory-theory proposed by Bartsch & Wellman (1995) treats mental states as theoretical entities, with the rules connecting these entities being acquired, learned and then used to make inferences when children grow up, giving rise to theory-of-mind. According to this theory, children make sense of other minds by learning and applying a general theory of the way minds work. Such understanding of mind develops through a succession of three, progressively more accurate theories, namely desire theory, desire-belief theory, and adult-like belief-desire theory. For very young children (those smaller than 2 years old), their theory of mind follows the desire theory, meaning that the children understand desires only and lack any understanding of belief. As they grow up, their desire-belief theory is developed and they begin to

understand belief. Yet, they only have a peripheral understanding of belief, which is easily overshadowed by desire. It is not until after 4 years old that an adult-like belief-desire theory is developed, which includes full integration of desire and belief for the successful performance of false belief tasks. This theory depicts theory of mind as an acquired capability that relies on a theory-mediated inference. In order to detect other people's mental states, the theory-mediated inference can draw on perceptually available information about the action taken by the target, other information in the environment as well as memory about the target's past behavior to infer the mental state. Whether or not such theory is stored in a specific site is still a matter of debate. Some authors (e.g. Fodor, 1992) view it as a special type of knowledge which is stored in a specific mental module, while others (e.g. Gopnik & Wellman, 1994) conceive it as just another kind of knowledge that is parallel to various types of world knowledge.

Adult theory of mind

Although general consensus from developmental research stipulates that theory of mind is maturely developed at around age 6 (Kesar, Lin, & Barr, 2003), such that adults are able to distinguish between external reality and belief held by others, people's childhood egocentrism has not been completely removed. Many social judgments have been found to be egocentrically based, and people tend to overestimate the extent to which others share the perception of the world as they do. For instance, people tend to believe that their mental states are more transparent to others than they actually are (Gilovich, Savitsky, & Medvec, 1998) and exaggerate the extent to which others will share their thoughts and feelings (Keysar, 1994; Van Boven, Dunning, & Loewenstein, 2000). Despite so, in the current literature on theory

of mind, there have been few studies investigating how adults deploy their theory of mind capacity.

Keysar, Lin and Barr (2003) argued that while adults have the ability to interpret social actions by means of a theory of mind, they do not deploy such ability reliably in interpreting others' behavior. In their study, participants played a referential communication game in pairs. One of the participants hid a roll of tape in an opaque paper bag, with the awareness that the other participant (the director) did not know what was in the paper bag. Yet, when the director asked the participant to "move the tape", referring to a cassette tape box which was mutually visible to both participants, 71% of participants attempted to move the paper bag in at least one out of four trails, and 46% attempted to move it for at least half of the trials, indicating that they did not take into account the information that the director was ignorant of the identity of the object in the bag. Their study provided evidence to the argument that though adults have a sophisticated theory of mind, it is not automatically deployed in understanding other's action and intention. Other contextual factors are at work.

Another study by Apperly, Riggs, Simpson, Chiavarino and Samson (2006) also points to the argument that adult's belief-reasoning is not automatic. The authors used an incidental false-belief task to examine whether people make automatic inference of other people's mental states. In their study, participants were shown video stimuli depicting actor placing and switching the location of an object, with an aim of indicating the location of the object at the end of each trial. At first, an actor hid an object in one of two boxes. Another actor then looked in the boxes and marked the location of the object on one of the boxes. In the experimental condition, after the second actor left the room, the first actor switched the boxes, resulting in a false belief in the second actor's part. Unexpected probe sentences tapping belief ("She thinks

that it's in the box on the right") or reality ("It's true that it's in the box on the right") were then presented for the participants to judge their correctness. The rationale is that as subjects needed to process information regarding the object's location in order to point out its location at the end of each trial, reality probe could be answered with information that had already been processed. On the other hand, if belief reasoning is automatic, information related to the second actor's belief would be updated automatically when the location of the boxes has been switched, so that reaction time to belief probes would be the same as that of reality probes as both are answered with information that has already been processed. It was shown that subjects responded more slowly to unexpected questions concerning another person's belief about an object's location than to questions concerning the object's real location. This difference in reaction time was not due to inherently different response time associated with belief and reality questions, as subjects showed no difference in response time to belief and reality questions when they were instructed to track the person's beliefs about the object's location.

All these investigations into adult's theory of mind suggest that adults do not reliably and automatically exhibit the mature theory of mind that is ascribed to them. Rather, the mature theory of mind capability in adults is deployed in a strategic way, such that we infer others' mental state only when it is called for in some specific situations. Indeed, such pattern of theory of mind deployment is similar to what developmental researchers identified from young children's understanding of mental states, where young children's performance on false belief tasks was significantly improved in specific contexts such as pretense and trickery. The above evidence thus contradicts the ToMM hypothesis, which argues that theory of mind is based on

domain-specific cognitive mechanisms that are fast, informationally encapsulated, and automatic (Apperly et al., 2006).

Study 1

In view of the mixed evidence regarding the automaticity of theory of mind, our study aims to further investigate the question of whether adults do automatically take into account other's mental states in understanding behavior. Indeed, although automatic cognitive processes have been proposed as the backbone of the theory of mind mechanism, little direct evidence for it has been reported. The present research extends Wertz and German's (2007) study in further investigating the automaticity argument, and because of this, a pilot study with an aim of replicating Wertz and German's (2007) study and finding out if any improvement in the procedure was needed had been conducted. Similar to the original study conducted by Wertz and German (2007), the hypothesis of this pilot study is that people would be more likely to incorrectly endorse a desire explanation for a distracter object in a false belief context, when the search for the target object led to the approach to that distracter object.

Participants

Five Cantonese-speaking undergraduate students from the Chinese University of Hong Kong participated in this pilot study for course credit.

Design and materials

The experimental stimuli used in Wertz and German's (2007) study was adopted. Short stories depicting a change in location scenario were shown to the participants on a computer screen, such as the following:

"Mary puts her hairdryer next to her perfume in the drawer and leaves the room. While Mary is in the shower, Gina moves the hairdryer to the cabinet. Mary comes back into the room for her hairdryer. She goes directly to the drawer."

In this story, the focal character, Mary, has a false belief about the target object (hairdryer). Because of this false belief, she approaches a location with a distracter object in an attempt to retrieve the target object (approach towards distracter, ADO). The other experimental condition involves stories depicting that the focal character approaches an empty location (approach towards empty location, AE):

"Mary puts her hairdryer in the drawer and leaves the room. While Mary is in the shower, Gina moves the hairdryer next to her perfume in the cabinet. Mary comes back into the room for her hairdryer. She goes directly to the drawer."

A question asking why the story character goes to a particular place ("Why does Mary go there?") was then presented after each story, followed by a sentence that either describes the belief (true or false) or desire (towards target or towards distracter) held by the story character, as shown in the following table:

Explanation type	**Example explanation**	**Correctness**
Desire for target	Because she wanted to get her hairdryer from the drawer	Correct explanation.
Desire for distracter	Because she wanted to get her perfume from the drawer	Incorrect explanation
False belief about target	Because she thought her hairdryer was in the drawer	Correct explanation.
True belief about target	Because she thought her hairdryer was in the cabinet	Incorrect explanation

Participants' task was to judge whether the sentence shown could correctly answer the question, and response was given by pressing one of the two response keys.

The stories, which had been translated to Chinese, were presented on the computer screen through E-prime, an experiment operating software. Each of the participants read and responded to 27 ADO stories and 27 AE stories presented in random order. Accuracy data as well as reaction time data were collected. We had also collected participants' qualitative feedback on the experimental procedure, especially probing whether they used any short-cut strategies in completing the experimental task, for possible improvement in the experimental procedure.

Results

With the small sample size of 5 participants in this pilot study, quantitative analysis did not yield any significant results. However, it was noted that participants completed the study within a very short period of time, on average taking less than 15 minutes in finishing the 54 trials. Qualitative inputs from the participants suggested that they used mental strategies to perform the task required in the experiment, the most common one being viewing and memorizing only the location and the object concerned without reading the whole story after the first few trials. In view of that, filler trials were included in study 3 to prevent participants from using mental strategies to complete the experimental task. The filler trials consist of stories similar to the experimental trials. But instead of probing the reason for the story character to approach a particular location ("Why does Mary go there?"), the filler trials probe for the location where the object desired by the story character was placed ("Where is the object wanted by Mary at last?"). To answer this filler trial correctly, participants need to read and understand the whole story, and the mental strategy reported by the participants in the pilot study would not work for both ADO and AE condition. Thus

it serves to prevent participants from memorizing only the key objects / locations and skipping the story.

Study 2

Although Wertz and German (2007) interpreted the result of their study as supporting the automaticity argument, an alternative explanation arguing otherwise exists. The error of attributing desire to the protagonist when s/he approaches the distracter in a false belief context could be a result of the strategic (non-automatic) nature of theory of mind, such that information about the protagonist's belief was not automatically taken into account when participants reasoned about the protagonist's desire. In the second study, I seek to critically test these two alternative explanations by presenting the participants with both false belief and true belief stories. In false belief stories, story character has a false belief about the target object, thinking that it is hidden in one location yet it has actually been moved to another location. When the story character searches for the target object in the original location, s/he then approaches a distracter object (in ADO story type) or an empty location (in AE story type). In such false belief stories, desire would not be helpful in interpreting the protagonist's action towards the distracter, as the search was purely directed by a false belief held about the target object, instead of a desire for the distracter object. In other words, the story character's action in false belief stories is driven by 1) a desire for the target, and 2) a false belief about the target, but not a desire for the distracter object. Contrarily for true belief stories, the story character is depicted as having a true belief about two objects placed in two separate locations. With a true belief about the location of both objects, desire for the object stored in a particular location would be the only plausible explanation for the story character's action towards that location.

Two different combinations of true belief and false belief stories, one dominated by true belief stories (in 'true belief stories dominant' condition) and the other by false belief stories (in 'false belief stories dominant' condition), were presented to two separate groups of participants. If theory of mind reasoning is automatically deployed, the performance of participants should be the same irrespective of the proportion that the two types of story are presented:

Hypothesis 1: Accuracy rate in rejecting the desire for distracter explanation in ADO stories does not show significant difference between 'true belief stories dominant' group and 'false belief stories dominant' group.

However, if theory of mind deployment is strategic in nature, participants shall be sensitive to the proportion of true belief to false belief stories, as these two types of story call for vastly different mental states in interpretation. Participants would then be inclined to use desire as explanation when the majority of stories presented are true belief stories, while belief would be more frequently used as explanation when the majority is false belief stories.

Hypothesis 2: Accuracy rate in rejecting the desire for distracter explanation in ADO stories would be lower for 'true belief stories dominant' group than 'false belief stories dominant' group.

Participants

Forty-one undergraduate students from the Chinese University of Hong Kong participated in this study for either course credit or small monetary reward. They were randomly assigned to one of two experimental conditions, namely 'true belief stories dominant' (N=18) and 'false belief stories dominant' (N=23). The proportion of males to females in each condition was approximately equivalent.

Design and materials

In order to test whether the attribution error observed in Wertz and German's (2007) study was due to the automaticity of theory of mind, apart from the false belief stories in Wertz and German's original study, true belief stories which require desire instead of belief as the underlying explanation for the protagonist's action were presented to the participants as well:

"Mary puts her hairdryer next to her perfume in the drawer. While Mary *is in the room, she watches* Gina move the hairdryer to the cabinet. Mary *wants to get something*. She goes directly to the drawer."

"Mary puts her hairdryer next to her perfume in the drawer. While Mary *is in the room, she watches* Gina move the hairdryer to the cabinet. Mary *wants to get something*. She goes directly to the cabinet."

In these two stories, Mary has a true belief about both the location of hairdryer and perfume. The correct explanation for why Mary goes to the drawer is that she wants the hairdryer, while that for the action of going to the cabinet is that she wants the perfume. In other words, true belief stories call for a desire explanation for protagonists' behaviour.

A question asking why the story character goes to a particular place ("Why does Mary go there?") was presented after each story, followed by a sentence that either describes the belief (true or false) or desire (towards approached-object or towards not-approached-object) held by the story character, as shown in the following table for the first true belief story quoted above:

Explanation type	**Example explanation**	**Correctness**

Desire for approached-object	Because she wanted to get her perfume from the drawer	Correct explanation.
Desire for not-approached-object	Because she wanted to get her hairdryer from the drawer	Incorrect explanation
False belief not-approached-object	Because she thought her hairdryer was in the drawer	Incorrect explanation.
True belief about approached-object	Because she thought her hairdryer was in the cabinet	Correct explanation

The stories in Chinese were presented on the computer screen through E-prime, an experiment operating software. Participants were instructed to read and understand the stories, after which a question asking why the story character goes to a particular place was presented. Following the question, sentence that either describes the belief (true or false) or desire (towards target/approached-object or towards the distracter/not-approached-object) was presented, and the participants' task was to judge whether the sentences presented could adequately and correctly answer the preceding question. Accuracy as well as reaction time data of the responses were collected.

There were 2 experimental conditions in this study with varying proportion of true belief versus false belief stories. As the response in the false belief trials is the focus of our analysis, its number across the two conditions was made constant. In both experimental conditions, participants were exposed to 32 false belief story trials. However, one group of participants (those in 'true belief stories dominant' condition) were presented with 68 true belief stories mixed with the false belief stories, while the other group (those in 'false belief stories dominant' condition) were presented with only 8 true belief stories mixed with the false belief ones. The sequence of presentation of the experimental trials was randomized. If the proclivity for attributing desire in ADO story type is due to the automaticity of theory of mind, the accuracy rate in explaining the action of the story character for the two groups should be similar.

Yet, if such proclivity is the result of a strategic theory of mind that fail to take into account certain mental-state information in interpreting action, the accuracy rate should be lower for the mental state explanation that is called for in the less frequently-occurring type of story.

The experiment lasted around 45 minutes for the 'false belief stories dominant' group and around 1 hour for the 'true belief stories dominant' group.

In analyzing the results, trials with story reading time shorter than 6 seconds were discarded. The data for 1 participant in 'true belief stories dominant' condition and that for 3 participants in 'false belief stories dominant' condition were discarded as their story reading time was shorter than 6 seconds in majority of the trials.

Results

Participants' accuracy rates for each explanation type in each story type, arranged by condition, were consolidated in Table 1. A 2 (condition; 'true belief stories dominant', 'false belief stories dominant') by 2 (story type; ADO, AE) by 4 (explanation type; 'desire for target', 'desire for distracter', 'false belief about target', 'true belief about target') analysis of variance (ANOVA), with repeated measures for the 'story type' and 'explanation type' factor was run to analyze the data. There was a main effect of explanation type ($F_{(3, 33)} = 11.278$; $p < 0.01$) and a main effect of story type ($F_{(1, 35)} = 5.946$, $p < 0.01$), qualified by a condition by story type interaction ($F_{(3,35)} = 4.807$, $p<0.01$). There was also a marginally significant three-way interaction between condition, story type and explanation type ($F_{(3,33)} = 2.701$, $p = 0.061$), as well as a marginally significant two-way interaction between story type and condition ($F_{(3,33)} = 2.830$, $p = 0.053$), indicating that participants' accuracy rate shows different

patterns under different combinations of explanation types and story types across the two conditions.

Table 1. Means and SD in all conditions for study 2

	DT		DD		FB		TB	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
True belief stories dominant								
ADO	100%	0	66.18%[a]	0.2643	93.63%[b]	0.1198	78.92%	0.3106
AE	98.53%	0.0606	84.80%[a]	0.1593	100.00%[b]	0.0000	86.76%	0.2186
False belief stories dominant								
ADO	89.17%	0.2420	81.25%	0.3128	93.75%	0.2276	89.58%	0.175
AE	93.75%	0.2276	83.33%	0.2621	95.00%	0.2236	83.33%	0.2493

Note. Different superscripts for pairs that are different significantly at p<.05

To better understand how participants' accuracy rate differs across the two conditions, a 2 (condition; 'true belief stories dominant', 'false belief stories dominant') by 4 (explanation type; 'desire for target', 'desire for distracter', 'false belief about target', 'true belief about target') ANOVA was run for each condition separately. For the true belief stories dominant condition, significant main effect for explanation type ($F_{(3,14)} = 14.25$; $p < 0.01$) and story type ($F_{(1,16)} = 9.33$, $p < 0.01$) were found, which was qualified by an interaction effect between story type and explanation type ($F_{(3,14)} = 3.63$, $p < 0.05$). Paired-sample t-test showed that there was significant difference in accuracy rate for 'desire for distracter' probe in ADO story and AE story ($t_{(16)} = 2.681$, $p < 0.05$), with accuracy rate in ADO story (M = 66.18%) lower than that in AE story (M = 84.8%). It was also found that the accuracy rate for false belief explanation type was significantly lower in ADO story (M = 93.63%) than AE story (M = 100%), $t_{(16)} = 2.193$, $p < 0.05$.

On the other hand, for the 'false belief dominant' condition, there was neither significant main effect nor interaction effect between explanation type and story type,

indicating that participants' accuracy rate across different stories and explanation types was similar.

Discussion

Participants presented with majority of true belief stories mixed with false belief stories showed greater error in mental state attribution, as they were more inclined to endorse desire explanations for the distracter object when the search resulted in an approach towards that distracter in a false-belief situation. However, for participants presented with majority of false belief stories, such error was not detected. This pattern is in line with the strategic hypothesis. In true belief stories where story character knows the location of the two objects correctly, desire would be the mental state accountable for the character's action. As the participant was exposed to true belief stories in most of the trials, they were accustomed to attribute story character's action as a cause of their desire. Thus, when false belief stories popped up, such strategic tendency led to the incorrect attribution of story character's action towards the distracter object as a desire towards that distracter object, instead of attributing it to the false belief held about the target object. This indicated the strategic nature of mental states attribution, which could be affected by the predominant mental state that the environment calls for in general, instead of what is particularly entailed in a single incident.

Study 3

Hypotheses

Although study 2 suggests that people's inference of belief or desire when we interpret other's action is strategic in nature, an important piece of information

associated with such interpretation is missing. The association between people's action and their mental state differs under different contexts because of situational demand. For example, if we saw that Mary read books at her leisure time, we tend to reason it in terms of her desire in reading - that Mary wanted to read books. Yet, if we saw that Mary read books at her mother's presence, then using desire in explaining her behavior may not be the most optimal due to situational demand. Mary might have done the action to please her mother. Thus, in different situations, the relevance of mental states in driving one's behavior differs. In this study we aim at investigating the effect of contextual cues in the deployment of theory of mind in interpreting action.

Given that mental states inference is dependent on situational cues, certain mechanisms, either incorporated within the theory of mind mechanism or separated from it, shall take into account such cues when we reason or make inferences about other's action. In low-demand situations, people's mental state is the main driving force of their action, and such situational cues shall increase people's alertness towards mental states. On the contrary, mental states are less relevant in explaining people's behavior in high-demand situations, rendering them less salient in such situations. Past researches have shown that salience of mental states was a factor affecting children's performance in false belief tasks (Chandler, Fritz, & Hala, 1989) and such saliency could be brought about by various situational manipulations. The most direct way would be highlighting the initial mental state of the protagonist in the experiment materials, such as mentioning the protagonist's desire or belief in the story. Thus, our first hypothesis is that if information about the story character's desire (in 'desire cue' condition) or motivation was provided, deployment of theory of mind

reasoning would be enhanced, and thus the error made in wrongly endorsing the desire for distracter explanation in ADO stories would be reduced.

Hypothesis 3: Accuracy rate in rejecting the desire for distracter explanation in ADO stories would be higher when the protagonist's initial desire for the target is mentioned, compared with control condition

Another way to manipulate the saliency of mental state would be by framing the task in terms of trickery. There has been abundant evidence supporting that mentioning deception enhances children's performance in false belief tasks. Several studies gathered evidence that framing the task in terms of explicit deceit reduces young children's egocentric error (Chandler, Fritz, & Hala, 1989), and the meta-analysis conducted by Wellman, Cross and Watson (2001) further supported that when a deceptive motive was explicitly stated, the performance in the false belief task was enhanced for children of all age. It was argued that framing the story in terms of trickery enhances children's performance through making mental state more salient in comparison with the alternative real state of affairs. Thus, our second hypothesis is that when information about deception is provided (in 'deception cue' condition), deployment of theory of mind reasoning would be enhanced, and thus the error made in wrongly endorsing the desire for distracter explanation in ADO stories would be reduced.

Hypothesis 4: Accuracy rate in rejecting the desire for distracter explanation in ADO stories would be higher when the story is framed in terms of trickery, compared with control condition

Participants

Sixty-two undergraduate students from the Chinese University of Hong Kong participated in this study for either course credit or a small monetary reward. They were randomly assigned to one of three experimental conditions. The proportion of males to females in each condition was approximately equivalent.

Design and materials

In studying how contextual cues affect adults' inference of mental states from other people's action, the methodology used in Wertz and German's (2007) study was adopted. Short stories depicting a change in location scenario, resulting in either approach to distracter or approach to empty location, were shown to the participants on a computer screen.

To study how contextual cues affect belief-desire reasoning, the stories were slightly modified to incorporate cues such as deception and original desire held by the character. To highlight the original desire held by the character, Mary, to the target object, the following sentence was added in the beginning of the original story:

"Mary loves her hairdryer"

To highlight the point that Mary went to the wrong location because she was deceived, the following sentence was used in the new story:

"Gina, in an attempt to fool Mary, moves the hairdryer next to her perfume in the cabinet"

The stories in Chinese were presented on the computer screen through E-prime, an experiment operating software. Participants were instructed to read and understand the stories, for example the following story highlighting the original desire held by the story character:

"Mary loves her hairdryer. Mary puts her hairdryer next to her perfume in the drawer and leaves the room. While Mary is in the shower, Gina moves the hairdryer to the cabinet. Mary comes back into the room for her hairdryer. She goes directly to the drawer."

After the story was presented, a question asking why the story character goes to a particular place was then presented:

"Why does Mary go there?"

Following the questions, a sentence was presented, which was selected randomly from the four types of candidate explanations as presented in the table below:

Explanation type	**Example explanation**	**Correctness**
Desire for target	Because she wanted to get her hairdryer from the drawer	Correct explanation.
Desire for distracter	Because she wanted to get her perfume from the drawer	Incorrect explanation
False belief about target	Because she thought her hairdryer was in the drawer	Correct explanation.
True belief about target	Because she thought her hairdryer was in the cabinet	Incorrect explanation

The participants' task was to judge whether the sentence presented could adequately and correctly answer the preceding question. Response was given by pressing one of the two response keys. Reaction time as well as accuracy of the responses were collected.

To prevent participants from using strategies to memorize the answers and skip the stories, filler questions asking the true location of the desired object (i.e. target) were presented intermittently at unpredictable intervals. In total, 90 stories, consisting of 54 test stories and 36 filler stories were presented to the participants. Given the large number of trials, the stories were divided into three blocks, with a 10

second resting time in between. The experiment lasted around an hour on average. Trials with story reading time shorter than 6 seconds were discarded.

Results

Participants' accuracy rates for each explanation type in each story type, arranged by condition, were consolidated in Table 2. A 3 (condition; control, deception cue, desire cue) by 2 (story type; ADO, AE) by 4 (explanation type; 'desire for target', 'desire for distracter', 'false belief about target', 'true belief about target') analysis of variance (ANOVA), with repeated measures for the 'story type' and 'explanation type' factor was run to analyze the results. There was a main effect of explanation type ($F_{(3, 57)} = 14.482$; $p < 0.01$), indicating that participant's accuracy rate differs for different explanation types. Other main effects and interaction effects were nonsignificant.

Table 2. Means and SD in all conditions for study 2

	DT		DD		FB		TB	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Control								
ADO	97.17%	0.0695	76.42%[a]	0.1828	94.98%	0.0751	94.61%	0.0894
AE	99.17%	0.0373	91.50%[a]	0.1130	96.33%	0.0613	92.89%	0.0876
Deception cue								
ADO	94.00%	0.1137	80.17%	17.45%	93.93%	0.0689	94.91%	0.0647
AE	96.50%	0.7210	79.75%	0.1953	91.51%	0.0992	96.61%	0.0491
Desire cue								
ADO	90.61%	0.1702	82.05%	0.2550	88.29%	0.1371	93.28%	0.1454
AE	94.55%	0.1503	81.89%	0.2414	88.22%	0.1817	92.12%	0.1758

Note. Different superscripts for pairs that are different significantly at p<.05

Given that our major focus in this study is on the accuracy rate of participant's desire attribution, a 3 (condition; control, deception cue, desire cue) by 2 (explanation

type; 'desire for target', 'desire for distracter') by 2 (story type; ADO, AE) analysis of variance (ANOVA), with repeated measures for 'story type' and 'explanation type' factor, was run for further analysis. Similar to the previous analysis involving both desire and belief explanation, explanation type main effect was noted ($F_{(1,59)} = 40.209$, $p < 0.01$). Apart from that, there was a story type main effect ($F_{(1,59)} = 4.663$, $p < 0.05$), indicating that participants' accuracy rate differs for the two story types (ADO and AE). Three-way interaction between explanation type, story type and condition was marginally significant ($F_{(2,59)} = 3.054$, $p = 0.055$), indicating that across the three conditions, participants showed different patterns in terms of their accuracy rate in different story types for different explanations.

Separate 2 (explanation type; 'desire for target', 'desire for distracter') by 2 (story type; ADO, AE) analysis of variance (ANOVA), with repeated measures for 'story type' and 'explanation type' factor was run for each of the three conditions to analyze the results further. For the control condition, significant main effects for explanation type ($F_{(1,19)} = 27.945$, $p < 0.01$) and story type ($F_{(1,19)} = 12.501$, $p < 0.01$) were found, which was qualified by an interaction effect between explanation type and story type ($F_{(1,19)} = 11.811$, $p < 0.01$). As revealed in paired-sample t-test, accuracy rate was significantly lower for 'desire for distracter' explanation type in ADO story (M = 76.42%), as compared with AE story (M = 91.50%) ($t_{(19)} = 3.823$, $p < 0.01$). There was no such pattern for the 'desire for target' explanation type. For deception cue condition and desire cue condition, although main effect for story type was found ($F_{(1,19)} = 26.379$, $p < 0.01$ for deception cue condition and $F_{(1,21)} = 5.139$, $p < 0.05$ for desire cue condition), there was no interaction effect between story type and explanation type ($F_{(1,19)} = 0.387$, $p = 0.541$ for deception cue condition; $F_{(1,21)} = 0.345$, p=0.563 for desire cue condition). The results indicated a greater difference in

accuracy rate in rejecting the 'desire towards distracter' explanation in the ADO story relative to the AE story in the control condition, as compared with deception cue and desire cue conditions.

Discussion

This study shows that contextual cues do affect people's performance in mental state attribution task. When the stories incorporated cues which are related to mental states (such as deception and original desire held by the character), participants were less prone to err in attributing the mental states responsible for story character's action. Participants were more reliable in rejecting the desire explanations for distracter object when the search was directed towards that object (as in the ADO story) when contextual cues were present in the story, as compared with the control condition where the contextual cues were absent. This suggests that in situation where mental states are more salient, people's attribution of mental states in accounting for other's action tends to be more accurate.

General discussion

The mechanism underlying theory of mind has been characterized as a fast, informationally encapsulated, and automatic cognitive mechanism under the ToMM model (Apperly et al., 2006). Yet, such characterization has been challenged by studies that pointed to the argument that adults' belief-reasoning is not automatic, as shown by the limitations on effective deployment of theory of mind (Keysar, Shuhong, & Barr, 2003). The current experiments attempt to test the automaticity of theory of mind by varying the saliency of the two key mental states (belief versus desire) when participants were in the process of making belief-desire reasoning, either through varying the proportion of experimental stimuli that call for different mental states in

interpretation (study 2), or through varying the content of the experimental stimuli and highlighting the mental states explicitly (study 3). Through this we also aimed to explore under what circumstances would the deployment of theory of mind be enhanced.

When the mental state of desire was frequently called for in attributing other's action in the frequent appearance of true belief stories, participants made more error in wrongly endorsing the desire explanation in attributing story character's action towards a distracter object, which was a result of having a false belief about another target object. According to Leslie and his colleagues (Friedman & Leslie, 2004; Leslie 2007; Thaiss, 1992), ToMM automatically parses behavior in terms of possible mental state descriptions, which were then reviewed by the selection processor, inhibiting the inappropriate mental states. Following this ToMM model, we would expect that participant's performance in the task would not be affected by the mix in true belief versus false belief stories, as separate possible mental state descriptions would be generated and reviewed automatically in each separate trial. However, we found that participants were more inclined to endorse the desire explanation for the story character's action in approaching the distracter object when they were exposed to more true belief stories, which calls for the mental state of desire in appropriately explaining one's action. This finding is incompatible with the ToMM model and suggests that adults' belief-desire reasoning is strategic in nature, which would be affected by the predominant mental states that has been adapted at a particular time.

Taking the view of a strategic nature of theory of mind, a question that follows is when, and what serves as a trigger for the strategic deployment of theory of mind. The most direct candidates would be situational cues that are related to people's mental states. The key question for my investigation is whether situational cues which

highlight the appropriate mental state for interpreting people's behavior would be picked up such that improvement in belief-desire reasoning would be shown. In study 3, it was demonstrated that when contextual cues relating to mental states (stories depicting a trickery situation or highlighting character's initial desire towards the target object) were presented, people were less prone to endorse the explanation that the character approaches the distracter object due to a desire towards that distracter object. This suggests that situational cues that enhance the saliency of mental states (which is desire in our case) serve as a trigger for the active deployment of theory of mind in interpreting other's action, thus resulting in a more accurate attribution. This is in line with German et al's functional MRI study (2004) showing that specific brain regions associated with explicit mental states attribution are automatically activated when participants view videos depicting pretense. It is possible that such cues were picked up automatically, activating the theory of mind mechanism and thus enhancing our performance of the task later on. The current findings also echo with evidence from infants' early belief attribution, where framing the task in terms of explicit deceit reduces young children's egocentric error (Chandler, Fritz, & Hala, 1989). It appears that the role of such situational cues in assisting the effective deployment of theory of mind is as important for infant as for adults.

In their original study conducted in 2007, Wertz and German argued that the automatic nature of theory of mind led participants to generate metal representation about the story character's desire when they approach the distracter object, thus resulting in the increased error rate in rejecting this 'desire towards distracter' explanation. However, such error also highlighted that participants did not automatically take into account story character's false belief in their belief-desire reasoning process. It pointed to dissociation between the mechanism underlying desire

attribution and belief attribution – while desire attribution is automatic, belief attribution is not. However, in this study, it was shown that even such desire attribution is not automatic, as it would be affected by the predominant mental state that the environment calls for, as well as the presence of cues that enhance saliency of mental states. It pointed to the possibility that when people are strategically engaged by the environment to employ their theory of mind capacity, they would perform better in inferring the belief and desire of others, thus making the right attribution.

With the findings in this thesis addressing that situational cues do play a role in enhancing people's performance in correctly attributing others' action to their desire, further studies could be conducted to investigate whether this could be extended to belief-attribution, or other mental states attribution. Theory of mind is a very broad topic encompassing various mental states which are relevant in understanding others and interpreting their action. Past studies have shown that development of our capability in understanding these mental states, and its subsequent performance are not necessary the same (Nguyen & Frye, 1999; Wellman & Woolley, 1990). In interpreting the results and making inference on the mechanism underlying theory-of-mind, we have to be cautious that the same mechanism may or may not apply to the understanding / reasoning of various mental states under theory-of-mind, and further studies would be needed to investigate whether different mechanism exists for different metal state reasoning.

Moreover, it would be beneficial to further investigate the mechanism under which situational cues affect our belief-desire reasoning. In particular, whether these cues are components of a broader theory of mind mechanism, or they just serve to initiate the active deployment of the theory of mind for desire-attribution. This would shed light on our understanding on a more comprehensive theory of mind model.

In evaluating the studies reported in this thesis, it was noted that some of the major findings are only marginally significant under the statistical tests. With only around 20 participants per experimental condition due to resources constraint, the sample size is considered small. In future research, a larger sample size would be more optimal to confirm the findings from the present studies. Also, given that the studies reported are centered around belief-desire reasoning, which is only one aspect of theory of mind, future studies could be extended to investigate factors affecting other adult theory of mind performance, such as 'hindsight bias' (Bernstein, Atance, Loftus, & Meltzoff, 2004; Fishhoff, 1975), and limits on effective deployment of theory of mind (Keysar, Shuhong, & Barr, 2003). Studies of different modalities, like actual interpersonal interaction in Keysar, Shuhong & Bar's 2003 study, instead of the current story reading and responding, could be conducted to extend the findings to a setting closer to real-life situations.

References

Apperly, I.A., Riggs, K.J., Simpson, A., Chiavarino, C., & Samson, D. (2006). Is belief reasoning automatic? *Psychological Science, 17,* 841-844.

Astington, N.J. (2003). Sometimes necessary, never sufficient: false-belief understanding and social competence. In B. Rapacholi & V. Slaughter (Eds), *Individual differences in theory of mind. Macquarie monographs in cognitive science* (pp. 12-38). Hove, E., Sussex: Psychology Press.

Axelrod, R. (1987). The evolution of strategies in the iterated Prisoners' dilemma. In L. Davis (Ed.), *Genetic algorithms and simulated annealing* (pp. 32–41). Los Altos, CA: Morgan Kaufmann.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind.* Cambridge, MA: MIT Press.

Bartsch, K., & Wellman, H.M. (1995). *Children talk about the mind.* New York: Oxford University Press.

Bernstein, D. M., Atance, C., Loftus, G. R., & Meltzoff, A. (2004). We saw it all along: visual hindsight bias in children and adults. *Psychological Science, 15,* 264-267.

Chandler, M., Fritz, A.S., & Hala, S. (1989). Small scale deceit: Deception as a marker of 2-, 3-, and 4-year-olds early theories of mind. *Child Development, 60,* 1263-1277.

Cassidy, D.W. (1998). Preschoolers' use of desires to solve theory of mind problems in a pretense context. *Developmental Psychology, 34,* 503-511.

Davidson, D. (1963). Actions, reasons, and causes. *Journal of Philosophy, 60,* 685-700.

Fodor, J.A. (1992). A theory of the child's theory of mind. *Cognition, 44,* 283-296.

Findler, N.V. (1990). *Contributions to a computer-based theory of strategies.* New York: Springer-Verlag.

Fishhoff, B. (1975). Hindsight does not equal foresight: the effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance, 1,* 288-299.

Friedman, O., & Leslie, A.M. (2004). Mechanisms of belief-desire reasoning: Inhibition and bias. *Psychological Science, 15,* 547-552.

Frith, C. (1994). Theory of mind in schizophrenia. In A.S. David & J.C. Cutting (Eds), *The neuropsychology of Schizophrenia.* Hove: Erlbaum.

Frith, C.D. & Frith, U. (1999). Interacting minds: A biological basis. *Science, 286,* 1692–1695

Galinsky, A.D., Maddux, W.W., Gilin, D., & White, J.B. (2008). Why it pays to get inside the head of your opponent: The differential effects of perspective taking and empathy. *Psychological Science, 19,* 378-384.

Gergely, G., Nadasdy, Z., Csibra, G., & Biro, S. (1995). Taking the intentional stance at 12 months of age. *Cognition, 56,* 165-193.

German, T.P., Niehaus, J.L., Roarty, M.P., Giesbrecht, B., Miller, M.B. (2004). Neural correlates of detecting pretense: automatic engagement of the intentional stance under covert conditions. *Journal of Cognitive Neuroscience, 16,* 1805-1817.

Gilovich, T., Savitsky, K., & Medvec, V.H. (1998). The illusion of transparency: biased assessments of others' ability to read our emotional states. *Journal of Personality and Social Psychology, 75,* 332-346.

Gopnik, A. and Wellman, H. 1994. The Theory-Theory. In L. Hirschfeld and S. Gelman (eds.), *Mapping the Mind: Domain Specificity in Cognition and Culture*. New York: Cambridge University Press. Pp. 257-293.

Keysar, B. (1994). The illusory transparency of intention: linguistic perspective taking in text. *Cognitive Psychology, 23,* 165-208.

Keysar, B., Lin, S., & Barr, D.J. (2003). Limits on theory of mind use in adults. *Cognition, 89,* 25-41.

Kinderman, P., Dunbar, R., & Bentall, R.P. (1998). Theory-of-mind deficits and causal attributions. *British Journal of Psychology, 89,* 191-204.

Leslie, A.M., Friedman, O., & German, T. (2004). Core mechanisms in 'theory of mind.' *Trends in Cognitive Science, 12,* 528-533.

Lillard, A.S. & Flavell, J.H. (1992). Young children's understanding of different mental states. *Developmental Psychology, 28,* 626-634.

Lopes, L.L. (1976). Model-based decision and inference in stud poker. *Journal of Experimental Psychology, 105*, 217–239.

Meltzoff, A.N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology, 31,* 838-850.

Nguyen, L. & Frye, D. (1999). Children's theory of mind: understanding of desire, belief and emotion. Social Development, *8,* 70-92.

Paal, T., & Bereczkei, T. (2007). Adult theory of mind, cooperation, machiavellianism: the effect of mindreading on social relations. *Personality and Individual Differences, 43,* 541-551.

Premack, D. & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1,* 515-526.

Saxe, R., Carey, S., & Kanwisher, N. (2004). Understanding other minds: Linking developmental psychology and functional neuroimaging. *Annual Review of Psychology, 55,* 87-124.

Siegal, M., & Beattie, K. (1991). Where to look first for children's understanding of false beliefs. *Cognition, 38,* 1-12.

Sperber, D., & Wilson, D. (2002). Pragmatics, modularity and mindreading. *Mind & Language, 17,* 3-23.

Sullivan, K., & Winner, E. (1993). Three-year-old's understanding of mental states: The influence of trickery. Journal of Experimental Child Psychology, 56, 135-48.

Van Boven, L., Dunning, D., & Loewenstein, G. (2000). Egocentric empathy gaps between owners and buyers: Misperceptions of the endowment effect. *Journal of Personality and Social Psychology, 79,* 66-76.

Wellman, H.M. (1991). From desires to beliefs: Acquisition of a theory of mind. In A. Whiten (Ed.), *Natural theories of mind: Evolution, development and simulation of everyday mindreading* (pp. 19-38). Cambridge, Ma: Basil Blackwell.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72,* 655-684.

Wellman, H.M. & Woolley, J.D. (1990). From simple desires to ordinary beliefs: the early development of everyday psychology. *Cognition, 35,* 245-275.

Wertz, A.E. & German, T.C. (2007). Belief-desire reasoning in the explanation of behavior: Do actions speak louder than words? *Cognition, 105,* 184-194.

Wimmer, H. & Perner, J. (1983). Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13,* 103-128.